JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 106, NO. C2, PAGES 2387–2405, FEBRUARY 15, 2001

Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements

W. J. Emery and D. J. Baldwin
Colorado Center for Astrodynamics Research, University of Colorado, Boulder, Colorado

Peter Schlüssel,
European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)
Darmstadt, Germany

R.W. Reynolds
National Climate Data Center, National Environmental Satellite Data and Information Service, National Oceanic and Atmospheric Administration, Camp Spring, Maryland

Abstract. The present computation of sea surface temperature (SST) from infrared satellite measurements requires a coincident sample of in situ (drifting buoy and/or ship) SST measurements, to compute by regression the algorithmic coefficients for the infrared data. Ignoring the fundamental difference between satellite-measured "skin SST" and buoy/ship measured "bulk SST," we analyze past buoy and ship SST data to better evaluate the errors involved in the routine computation of SST from operational satellite data. We use buoy and ship SST data for 2 years (1990 and 1996) from the Comprehensive Ocean-Atmosphere Data Set as well as 2 years of previously cloud-cleared satellite radiances with matching drifting/moored buoy SST data from the NASA Pathfinder SST data set. We examine the in situ SST data for geographic distribution, accuracy, and self-consistency. We find that there are large geographic regions that are frequently not sampled by the present drifting buoy network, a natural consequence of the fact that most buoys are not deployed to measure in situ SST for satellite infrared SST calibration. Comparisons between drifting buoy SSTs suggest an error of ~ 0.4°C for nearly coincident buoy SSTs. Comparing moored with adjacent drifting buoy SSTs, we find that drifting and moored buoy SSTs are samples from the same population. Ship SSTs are noisier and have a significant warm bias relative to drifting buoy SSTs. We explore the SST measurement accuracy changes that occur with variations in sampling coverage used for the SST algorithm regression. We both vary the total amount of points and restrict the regression data to regional sampling biases. Surprisingly the total number of calibration SST values can be quite small if they cover all latitudes. We conclude that buoy SSTs can have residual bias errors of ~0.15°C with RMS errors closer to 0.5°C.

1. Introduction

The regression procedure used widely to compute algorithm coefficients for the computation of sea surface temperature (SST) from infrared (IR) satellite radiances inherently assumes that some selected subset of global drifter SSTs is a reference SST set without errors. This calibration step is needed because the satellite IR instruments (and their onboard calibration systems) drift over time and there is a variable atmosphere that affects the radiation passing between the satellite and the Earth's surface. Previous studies [*McClain et al*, 1983, 1985; *Walton et al.*, 1998] have simply corrected for these variations by assuming that the buoy-measured in situ SST can be used as representation of the "truth" that the satellite radiances can be regressed against.

While nobody would claim that the buoy SSTs are perfect, there have been very few studies that have focused on evaluating the in situ SSTs for accuracy, precision, self-consistency, and sampling representativeness. We examine global drifter and ship SST data for 1996 and 1990 along with satellite radiances that have been carefully produced by the NASA SST Pathfinder project [*Evans and Podesta*, 1996]. Thus, for our comparisons with satellite SSTs we do not have to do any additional "cloud clearing," and the satellite radiances can be assumed to be under clear-sky conditions as the radiances have been cloud cleared with a consensus algorithm. Thus we will assume that these data are as free of cloud contamination as is humanly possible.

Paper number 2000JC000246
0148-0227/01/200036000246$09000

2. Space-Time Distributions of In Situ SST Data

2.1. Annual Means

Before looking at the statistics of the buoy populations, we need to discuss what SST drifting buoys measure and how well. Because of vertical motion in the wave field the drifting buoys' SST sensors generally measure SST about 1-2 m beneath the sea surface (Niiler, personal communication, 1998; Swenson, personal communication, 1998). Many of the sensors on these buoys are initially calibrated to an accuracy of ±0.1°C before they are installed and once they are deployed there is no post-deployment SST calibration possible. Mark Swenson of the U.S. Drifting Buoy Data Center at the Atlantic Oceanographic Marine Laboratory (AOML) in Miami, Florida, reports that most of the present drifting buoys have their temperature sensors in the same position and that there are only three different types of buoy hulls. He believes that the buoy SSTs are accurate to about ±0.15°C. It is not possible to evaluate any operational lifetime sensor drift or change in the sensor calibration, and we must evaluate the buoy SST on the basis of their own values.

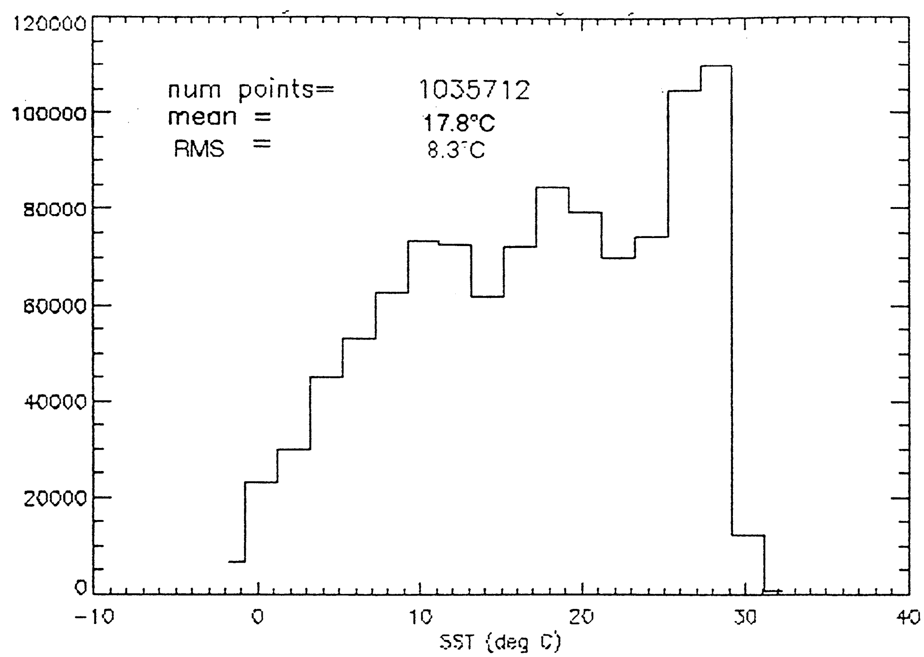

Figure 1a. Annual histogram for the global drifting buoy SST data for 1996.

Since we do not have detailed calibration information on each of the buoy SSTs nor do we have the capability of recalibrating any of these buoy SST sensors, we can only examine the data themselves to evaluate their character and consistency. First we want to determine just how and where these buoys are measuring SST. We must recognize that because of wave motion the drifting buoys measure the rough average temperature of this (1-2 m) upper layer. In addition, none of the drifting buoys was explicitly deployed to measure SST for the calibration-validation of satellite infrared measurements. Many investigators argue that their buoys will provide this SST calibration-validation information but their primary object is to provide a Lagrangian measure of the surface current. We selected a period of time when we have a uniform set of satellite infrared measurements for comparisons with the buoy data and have selected the global buoy SST measurements from 1996. The overall histogram of these SST data is presented here in Figure 1a. More than 1 million SST data values were available globally for this period. The highest peak value of this histogram is located at ~ 27°C with two smaller peaks at 18°C and 10°C. The overall mean of this distribution is ~

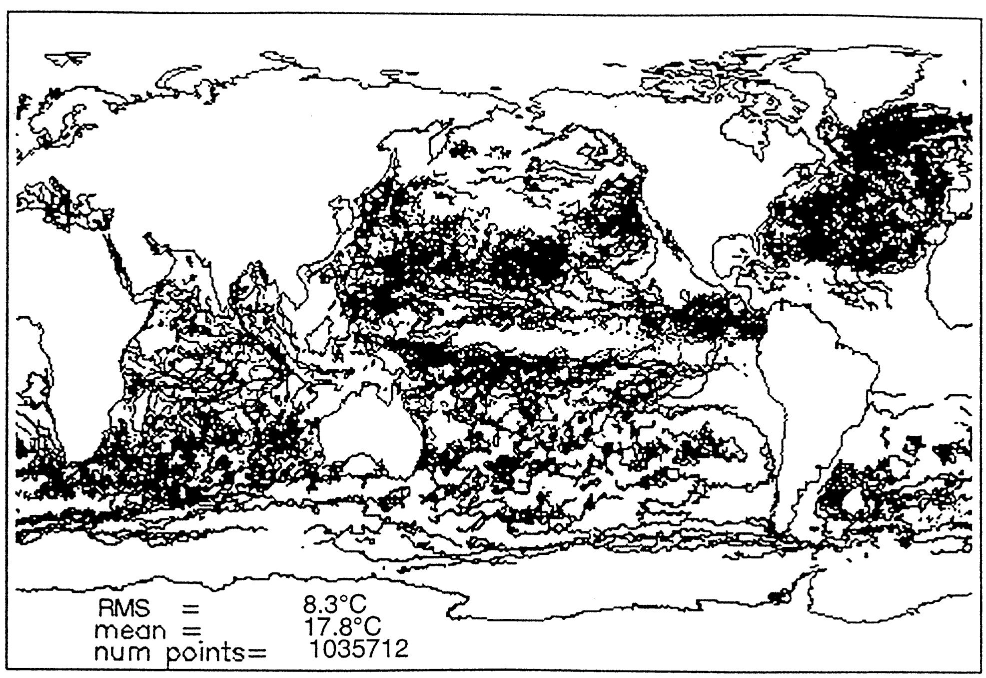

Figure 1b. Geographic distribution of drifting buoy SST observations for 1996.

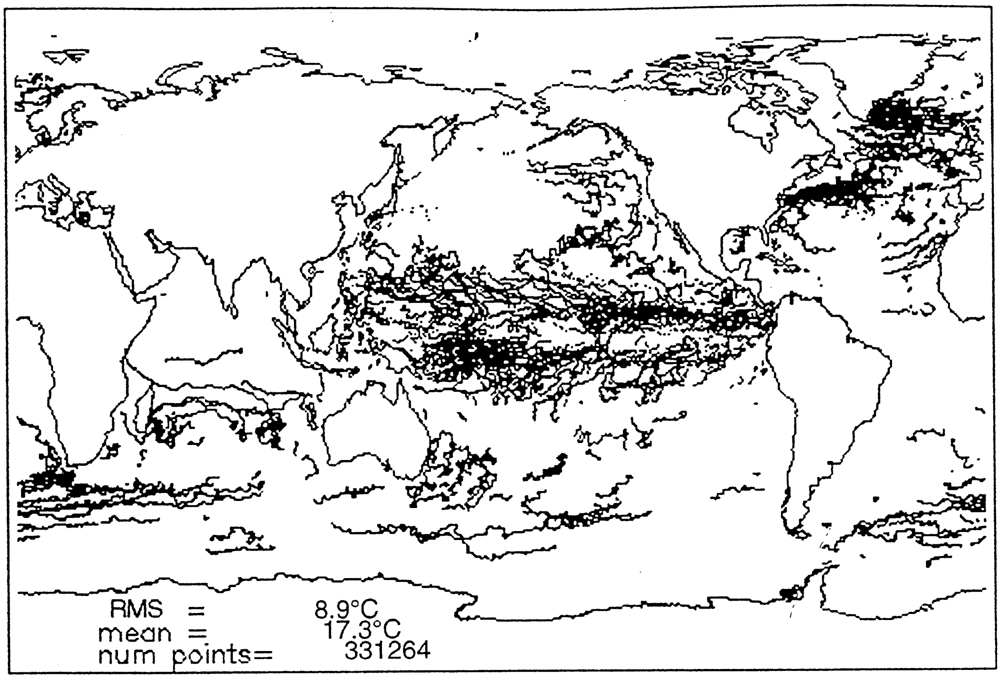

Figure 2a. Global buoy SST distribution for 1990.

18°C with an RMS difference of about 8°C. The general distribution is skewed to the higher values with a significant drop-off toward the colder values.

This histogram reflects the geographic distribution of the buoy SST data presented here in Figure 1b where the individual buoy SST sample locations are plotted for the entire time period. At first glance the global coverage is quite impressive, especially for this relatively short period of time. Closer inspection, however, reveals the lack of samples at the highest latitudes, particularly in the North Pacific and the Southern Ocean. There is an absence of buoy SST measurements in the equatorial Pacific because of upwelling, which is typical of all of the geographic distributions we have examined. The equatorial and the South Atlantic are also lacking in buoy SST coverage. There are other, smaller regions where no buoy SSTs have been collected in this period, such as the upwelling region west of South America or the coastal region east of Southeast Asia. The summary statistics for Figure 1 a have been displayed on the

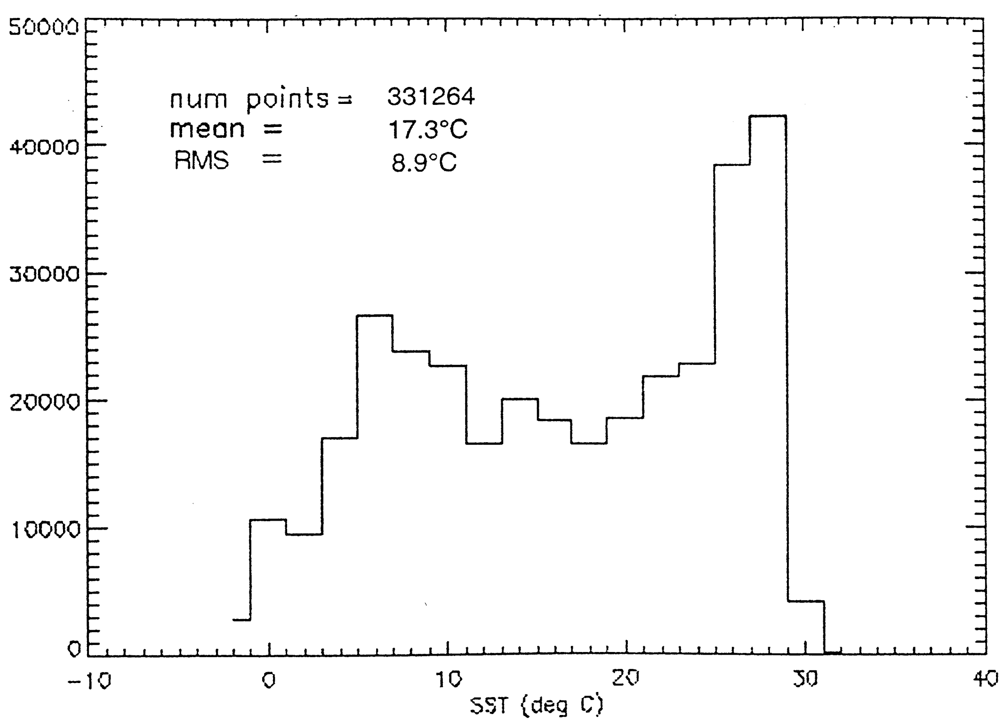

Figure 2b. Global histogram of the buoy SST data for 1990.

corresponding geographic distribution of Figure 1b just for reference. Once again, the reader is reminded that none of these buoy SST measurements were expressly made to provide calibration information for satellite infrared SST estimates.

A comparison with the annual mean buoy SST distribution for 1990 in Figure 2a reveals how superior the buoy coverage was in 1996. In 1990 the buoy data were almost exclusively restricted to the tropical Pacific and the northern North Atlantic. There are a few buoys in the northern parts of the Southern Ocean and a few in the eastern North Pacific. There are no buoys in the Indian Ocean, the midlatitude and western North Pacific, the South Pacific, and the tropical and South Atlantic. There are sections of the equatorial Pacific that are again absent of buoy SST observations. The coastal ocean off of southeast Asia is also unsampled.

This buoy SST coverage is much more restrictive than that for 1996, suggesting that SST algorithms computed from this set of buoy SSTs may not have included many of the important SST conditions that should be sampled for a comprehensive, global SST equation. The corresponding histogram in Figure 2b also indicates a more limited sample than was found in Figure 1a. The first striking difference between the 1990 histogram in Figure 2a and the 1996 histogram in Figure 1a is the difference in the overall number of buoy SST observations. The 1996 high-temperature maximum at 28°C has a value of well over 100,000 observations while the highest peak in Figure 2b, still at 28°C, is just over 40,000 observations. Thus, the 1990 global sample has less than half the data values available in 1996.

These far fewer observations in 1990 are also distributed quite differently. The 1990 histogram (Figure 2b) is clearly bimodal while that for 1996 is not. The lower temperature peak in Figure 2 is at ~ 6°C where there is no corresponding peak in the 1996 histogram in Figure 1a. This lower temperature peak in 1990 corresponds to the buoy SST samples from the high-latitude, or subarctic North Atlantic (Figure 2a). While there are SST values from the high-latitude North Atlantic in 1996 (Figure 1b) they are not dominant in the 1996 global buoy SST distribution. For example, there are over 40,000 observations at 6°C in the 1996 histogram (which is not a peak) while the clearly marked 6° C peak in the 1990 histogram has just under 30,000 observations. It is important to recognize that the buoy SST coverage is not globally uniform and that geographic biases may strongly influence the formulation of any SST algorithm dependent upon regression against drifting buoy SSTs.

2.2. Drifting Buoy SSTs: Seasonal Variations

The SST histogram for our 1996 Northern Hemisphere winter period (Figure 3) is different from the overall histogram in Figure 1a. There are now four peaks in the histogram with the largest peak at 22°C. These peaks are separated by one bin and are most likely not significant. However, between roughly 5°C and 29°C the distribution is more evenly distributed over temperature than it was in Figure 1a. This shift reflects the fact that more of the buoy measurements are in the Northern Hemisphere (Figure 1b), where the SST is colder in the winter leading to the lower histogram peak at 22°C. The overall mean is approximately 17°C with an RMS difference of 8°C, which is very close to the summary statistics of the annual histogram in Figure 1a.

The Northern Hemisphere spring histogram in Figure 3 has a mean and an RMS very similar to the annual mean and winter values. Unlike the winter, however, the tallest histogram peak is again at the higher temperature of ~ 28°C. The overall histogram shape for this season appears quite similar to the annual histogram in Figure 1a. The summer histogram shifts to a much larger

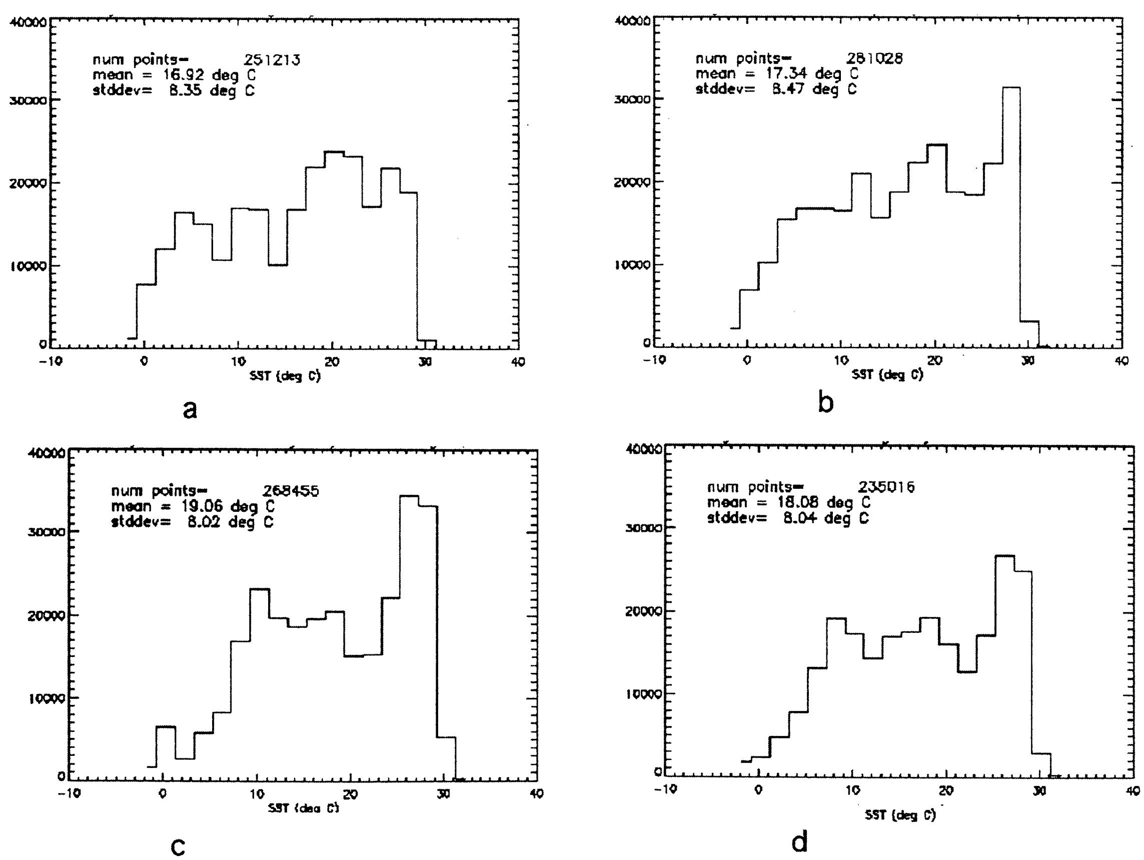

Figure 3. Histograms of drifting buoy SSTs for the four Northern Hemisphere seasons of 1996, (a) winter, (b) spring, (c) summer, (d) fall

maximum at ~28°C. This time there is a change in the mean value to just over 19°C with an RMS difference of ~8°C. Again, this reflects the fact that more of the SST samples are in the Northern Hemisphere, and thus a seasonal shift in the mean SST is expected that would result in the higher-temperature peak during the Northern Hemisphere summer. It is interesting that in this summer histogram there is a more dramatic falloff for temperatures below 10 °C which is not characteristic of the annual mean or spring histograms. The fall histogram shifts the mean value back to 18°C with an RMS of ~ 8°C. Here the 28°C peak still dominates the histogram with a shape roughly similar to that of spring.

We examined the seasonal geographic distributions of the buoy samples used to compute these seasonal histograms. As expected, there is little change in geographic distribution of these buoy SSTs, which all have an appearance similar to that in the annual distribution in Figure 1b. All of these maps exhibit the dominance of the Northern Hemisphere for the buoy SST coverage particularly in the Atlantic where the coverage in the South Atlantic is very meager.

2.3. Moored Buoy SSTs

The histogram of the moored buoys (not shown) is strongly bimodal which would be expected from the distribution of the buoy locations which are in the tropical Pacific and along the east and west coasts of the United States. The 27°C histogram peak is the expression of the Pacific equatorial SST measured by the Tropical Atmosphere-Ocean (TAO) buoys in the tropical Pacific. The long time series of data collected at these locations represents a unique set of SST measurements. The smaller peak at about 10°C represents the SST measurements made by the moored buoys located along the east and west coasts of North America, and off the coasts of Japan and western Europe (especially the United Kingdom). With the exception of the moored buoys near Hawaii there are no operational moored buoys located neither in the subtropical regions nor in many other parts of the world-ocean. Again it must be observed that these moored buoy SST measurements were made as part of other programs and that none of the moored buoys was expressly operated to provide SST measurements for satellite calibration. Furthermore, it should be noted that only the TAO buoys are used for regression against satellite SSTs [*Walton et al.*, 1998]. Other moored buoy observations were not used because they were assumed to be in high-gradient coastal regions which would require relatively close collaboration between satellite and buoy observations to eliminate spatial errors. It is surprising, however, to see how well the overall statistics from this very limited moored buoy SST sample compare with those from the drifting buoy SSTs. The moored buoy mean SST is just over 18.5°C which is less than a degree above the 17.9°C 1996 mean for the drifting buoys. At 8.6°C the moored buoy SST temperature RMS is only about 0.4 °C above that for the drifting buoys.

2.4. Ship SSTs

Turning to the corresponding (in time and date) global ship SST reports (Figure 4a) we find that there are just slightly more (1,132,427 versus 1,035,712) ship SST reports than for the corresponding drifting buoy SST data. The ship histogram is very different than those for the drifting and/or moored buoy SSTs in spite of how closely matched the overall statistics are from the histograms. The mean ship SST is 18.7°C just a degree above that for the drifting buoys (17.8°C) and is very similar to the 18.5°C mean for the moored buoy SSTs. The ship RMS SST of 8.4°C is very close to the 8.2°C drifting buoy SST RMS temperature difference and is identical to the moored buoy SST variability. The ship SST histogram has only a single large peak at about 27°C, which is twice as large as most of the other histogram values. There is a falloff with decreasing temperature below about 15°C which is very similar to the falloff in the annual mean drifting buoy histogram (Figure 1a) below about 10°C.

The geographic distribution (Figure 4b) of the ship SST samples explains many of the facets of the mean annual ship SST histogram. There is an extreme concentration of ship SST observations for the Northern Hemisphere with the North Atlantic being particularly filled with measurements even in the subarctic regions. Conversely, there are very few ship SST measurements in the Southern Ocean except for a few tracks of resupply ships. The Indian Ocean is poorly covered compared to the Northern Hemisphere ocean basins. Only a couple of ship sections cut across the central parts of the Indian Ocean. The eastern South Pacific is poorly covered with ship SST observations, as is the South Atlantic. The concentration of measurements in the tropics and the subtropics accounts for the 27°C major peak in the overall mean 1996 ship SST histogram. While there are a lot of measurements in the northern high latitudes there are none in the south, creating the decrease with temperature found in the overall 1996 mean ship SST histogram.

3. In Situ SST Consistency

A basic problem in evaluating the accuracy of in situ SST data used for satellite reference values is the lack of an independent measure of SST to act as a reference. In the comparisons, which follow in sections 3.1 – 3.4, we compare the different types of in situ data (drifting buoys, moored buoys and ships) with themselves and with each other. In these comparisons we compute mean, and RMS statistics. However, in the intercomparisons within the same type of observation (e.g., drifter with drifter SST) a combination of all possible pairs must have a zero mean because each pair will occur twice in the mean statistics with opposing sign. In this case, the standard deviation and RMS are identical, so for clarity we have elected to label all values as RMS. This helps to establish just how self-consistent these data are and gives us an estimate of the basic accuracy level of the in situ SST data.

3.1. Drifter Versus Drifter SST

To reduce the number of computations, we restricted these comparisons to four months representative of the seasonal cycle in 1996. For example, we present the results for March 1996 in Figure 5. As expected, the temperature differences increase with increasing separation. Still, there is considerable scatter at almost all temperatures. Even at the smallest separations the variability is significant. The overall RMS difference of the temperature differences is 0.78°C which is larger than expected.

As might be expected, this plot shows an overall decrease in the temperature difference as the separation distance decreases. At the wider separation both instrument noise and space-time geophysical variability contribute to this temperature difference. As the separation gets smaller the space-time variations are reduced, and the final temperature difference should be due primarily to instrument noise, thus representing a threshold value for the SST measurement.

To better evaluate the SST differences at the smallest separations, we have expanded this relationship for separations between 0 and 50 km (Figure 6). There is a substantial variability represented by the RMS difference which has a mean value of 0.69°C over this 50 km interval. Over this interval the RMS decreases from a high at the 50 km separation of ~1.0°C, to a minimum of ~0.15°C at the 10 km separation. Thus, even at the smaller separations there is an uncertainty of ~0.15 °C in any of these buoy SST estimates.

Another important observation is that this RMS temperature difference increases rapidly from its minimum at 5 km separation and that by the time the separation is ~25 km the temperature difference has increased to be >1.0°C. Thus in situ SSTs should not be matched with satellite radiances more than a few kilometers away. Note that in Figure 7 the number of observations (SST pairs) is constant at about 200 points at buoy separation greater than 10 km buoy while below this separation

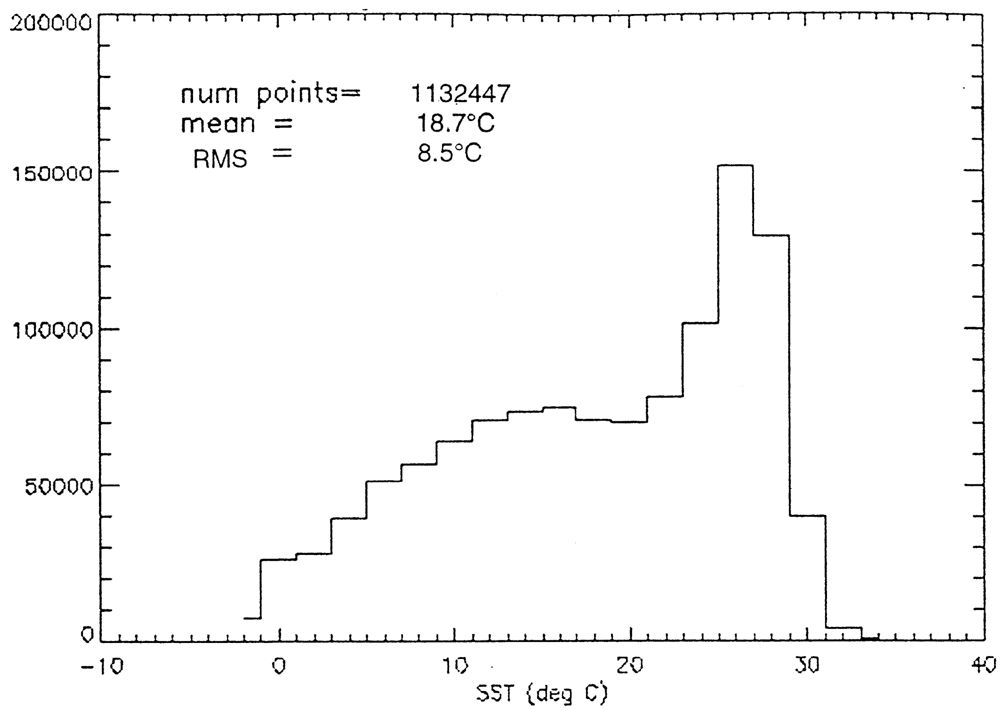

Figure 4a. Histogram of global ship SST data 1996.

distance the number of observations increases dramatically to be well over 1000 points.

In order to evaluate the possible seasonal variations in these buoy to buoy SST comparisons we computed the same statistics for July, October and January of 1996. Rather than repeat all of the figures for these different seasons we instead summarize the statistics in Table 1, where we give the overall RMS difference, the 0-50 km RMS differences, the RMS difference at 5 km and the total number of observations. The overall RMS differences are quite similar and range from 0.78 to 0.90°C. For the expanded 0-50 km interval the average RMS differences have reduced to about 0.5°C ranging now between 0.47° and 0.68°C. Finally, the 5 km RMS temperature differences go from 0.1° to 0.45°C. The number of observation pairs are between 8000 and 10,000. The change in variability at 5 km during the different months is affected by the low number of observation pairs at this short interval and is not significant. From Table 1 it seems safe to conclude that the buoy SST accuracies are in general not better

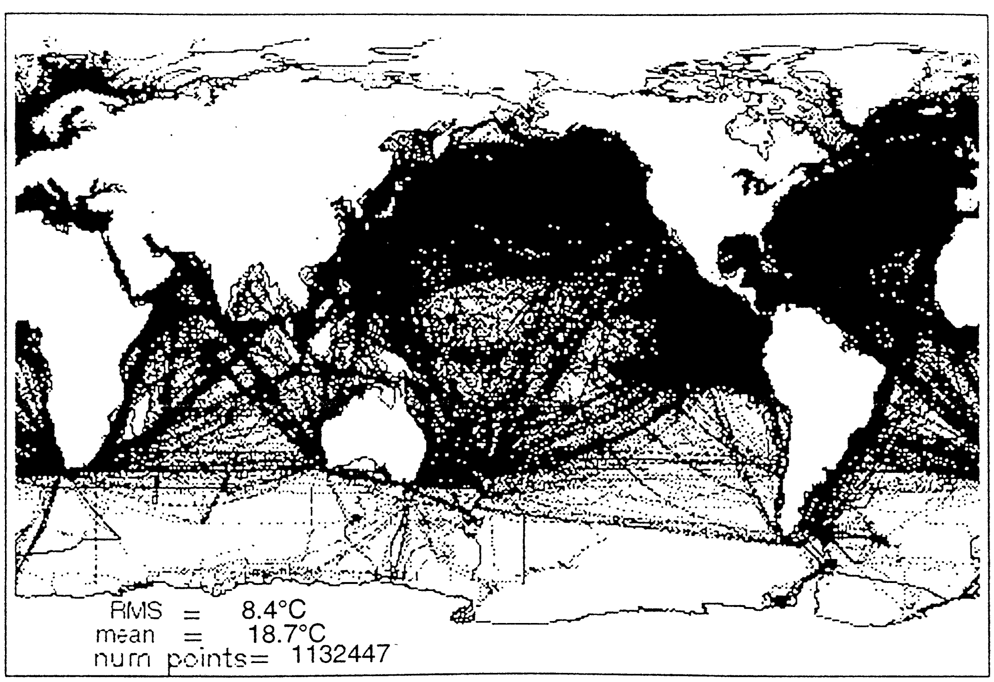

Figure 4b. Data distribution for ship SST in 1996.

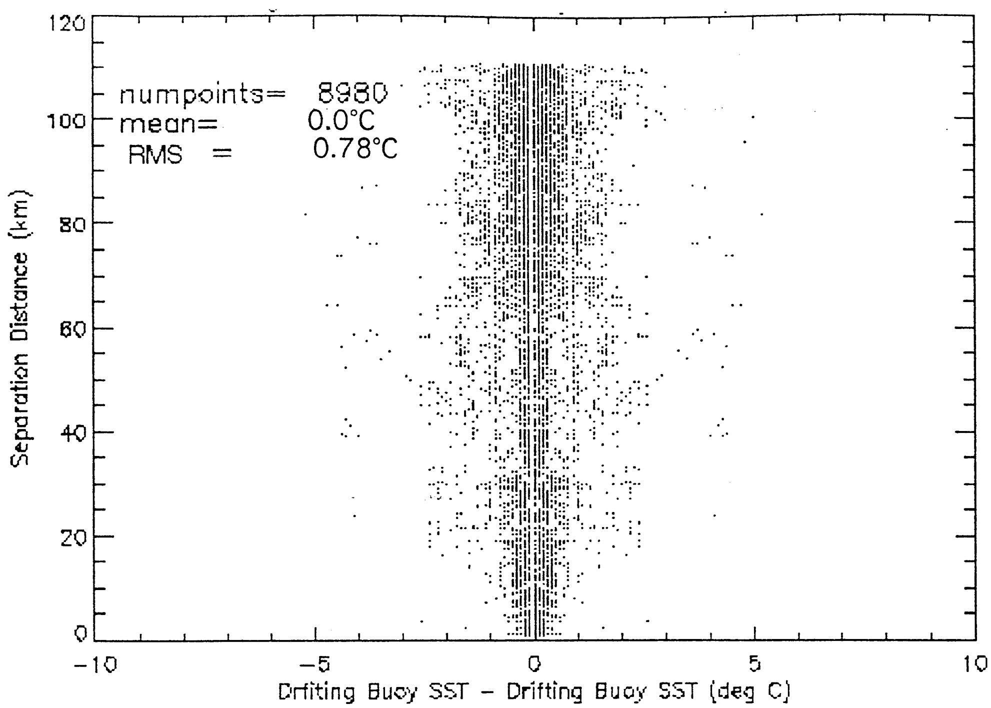

Figure 5. Plot of drifter versus drifter SST differences as a function of separation distance for March 1996.

than 0.3°C. In addition, the matchups between in situ and satellite temperatures should not be for separations greater than 5-10 km.

3.2. Comparison Between Drifting and Moored Buoy SSTs

It is usually assumed that drifting buoys and moored buoys measure the same temperature near the surface. As a consequence most studies have lumped these SST estimates together in producing statistics for the satellite SST community. We decided to use our in situ data set to compute this difference expressly. For the global data set (which does not include the equatorial Pacific moorings) the mean difference is -0.96°C which is a significant difference for the some 300 comparison points. All differences were computed as moored minus drifter SST, and thus

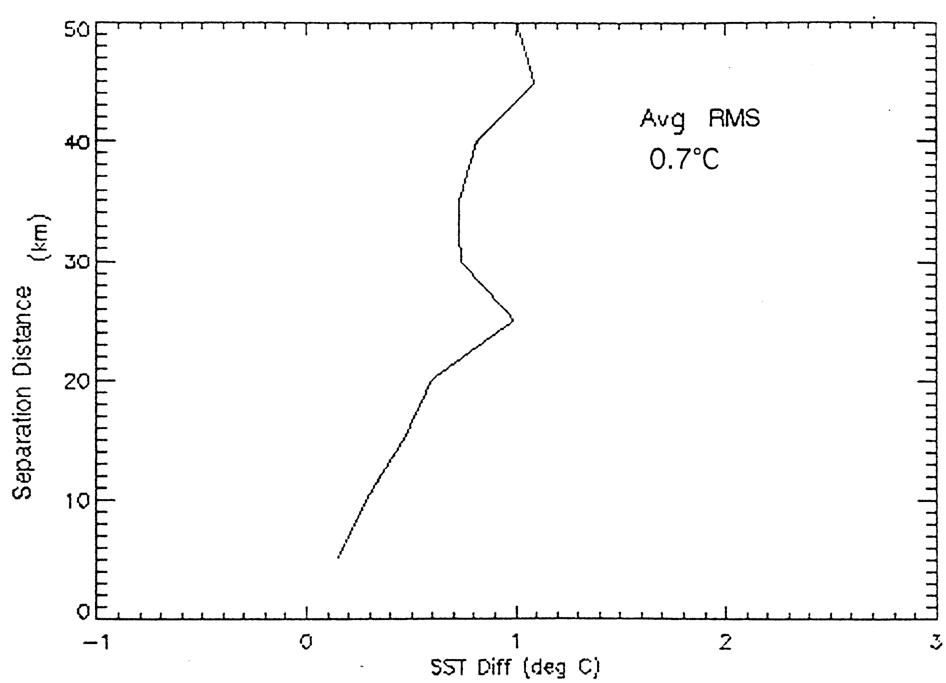

Figure 6. RMS buoy to buoy SST differences for 0-50 km for March 1999.

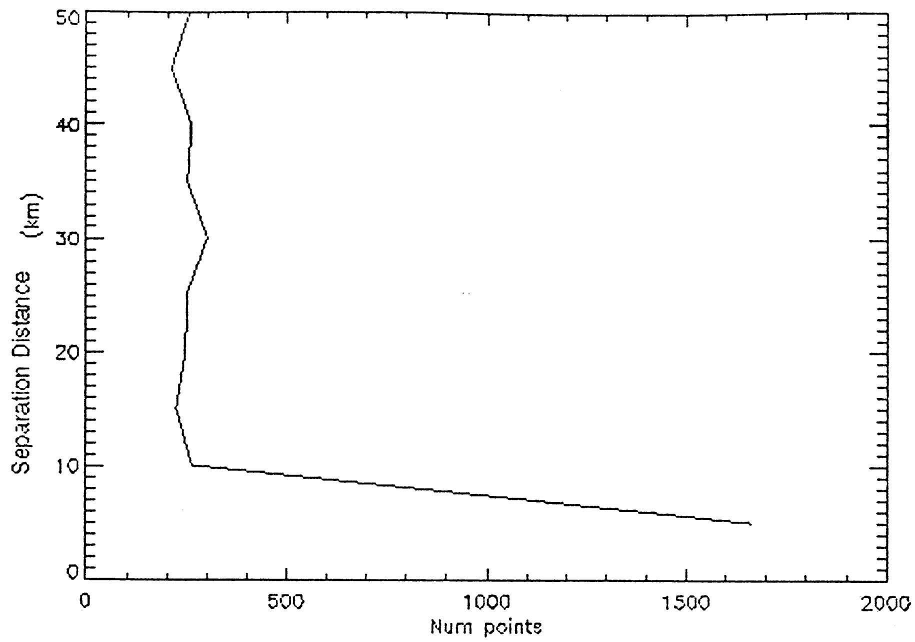

Figure 7. Number of observation pairs for 0-50 km separations of buoy to buoy SST differences for March 1996.

the minus sign indicates that the moored buoy SSTs are slightly cooler than the drifting buoy SSTs. A similar comparison carried out for the equatorial Pacific alone yielded a mean temperature difference of 0.05°C and an RMS difference of 0.1°C, indicating that the equatorial TAO moored buoys were considerably more like the coincident drifting buoy SSTs.

3.3. Ship Versus Drifting Buoy SSTs

Turning to a comparison between drifting buoy and ship SSTs, we present in Figure 8 the differences between the ship and drifting buoy SSTs again as a function of separation distance. The positive mean difference of 0.28°C is consistent with the observation that ship injection SSTs are slightly warmer due to the heating in the engine room where the observations are made [*Saur*, 1963]. The ship-buoy RMS difference at 1.8°C is about twice the size of the drifter versus drifter RMS difference and is also consistent with the fact that ship SSTs are found to be a lot noisier than buoy SSTs [*Kent et al.*,1993; *Kent et al.*, 1999; *Kent and Taylor*, 1997]. The visual reading of the injection SST and the recording of the ship SST introduce some of this variability by hand. The radio reporting of ship SSTs also introduces deviations that will be counted here as noise. Still, the overall distribution is consistent with the concept that SST differences will be larger with increasing separation. The variability of the SST appears as expected to increase with separation distance.

Expanding again the ship versus buoy SST differences at separations less than 50 km, we have the two curves in Figure 9. Unlike the buoy to buoy SST comparisons with their zero mean, the comparisons between the ship and buoy SSTs have a nonzero mean. The mean has a slight positive bias at ~0.2°C over this 50 km interval, again expressing slightly warmer ship SSTs. The mean RMS temperature difference at 1.4°C is just a bit more than twice that for the buoy versus buoy SST comparison. The variations in RMS temperature difference with decreasing buoy separation are clearly not significantly different, and only the mean RMS value of 1.4°C has any statistical significance. Again the abundant noise known to be in the ship SST data is the primary cause for this increase relative to the buoy-only comparisons. Even with this very noisy behavior, however, the RMS temperature differences decrease from a 50 km value of 1.4°C to ~ 0.9°C at the 5 km separation. Unlike the buoy to buoy comparisons the number of observations in Figure 10 increases almost linearly from about 10 points at 5 km to about 100 points at the 50 km separation. At the smaller separations there are not enough comparisons to yield a statistically significant mean temperature difference. This is best demonstrated in Table 2, which gives the number of observations separated by less than 5 km.

To assess the possible seasonal dependence of these comparison statistics, we again compute all of these temperature differences and present their statistics in Table 2. There is greater variability of the RMS temperature differences in Table 2 than was found in Table 1. The overall mean difference ranges from 1.54°C in July to a maximum of 193 °C in October. A similar range is found in the 50 average RMS differences, which go from a minimum in October of 1.12°C to a July maximum of 1.77°C exactly the reverse of the overall differences. This suggests that these variations are random and do not reflect any systematic seasonal

Table 1. 1996 Drifter-Drifter SST Statistics

Month	Overall RMS (°C)	50 km RMS(°C)	5 km RMS(°C)	Observed	Observed (<5km)
Jan.	0.90	0.47	0.10	8,250	1160
March	0.78	0.68	0.15	8,980	1675
July	0.81	0.51	0.45	10,274	2250
October	0.89	0.65	0.20	8,262	1805

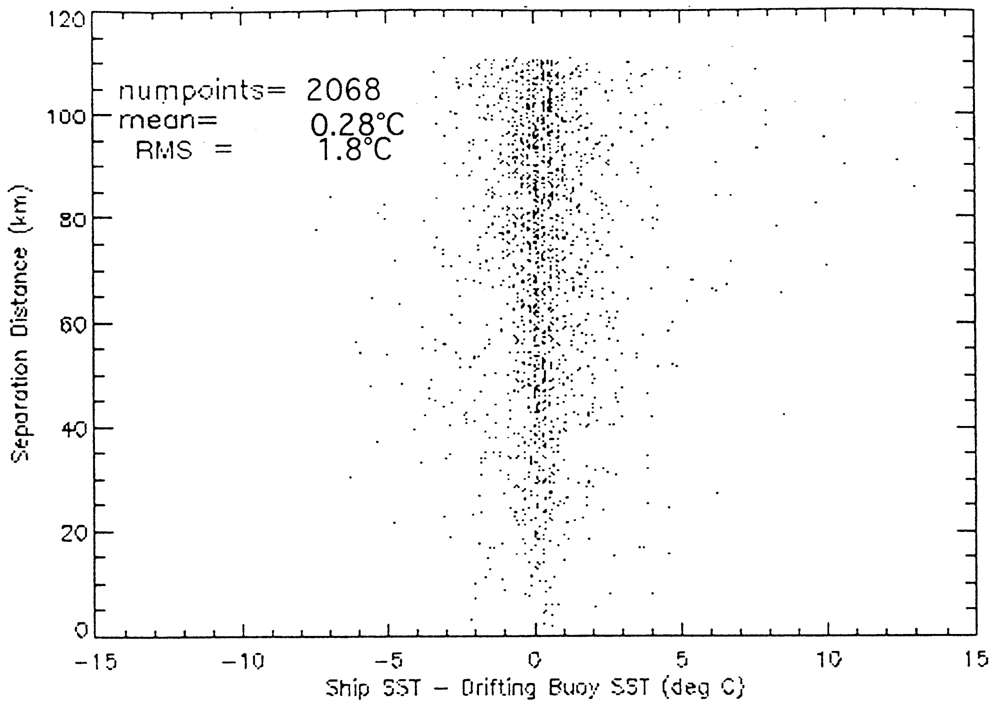

Figure 8. Ship versus buoy SST differences as a function of separation distance for March 1996.

changes. The RMS differences at the 5 km minimum separation are generally ~1.0°C which is much larger than the buoy-only comparisons. The number of observations is generally ~ 2000 points which is a value restricted by the buoy availability and not the ship SSTs.

3.4. Ship Versus Ship SSTs

The next comparison is between ship SSTs and other ship SSTs (Figure 11). The first thing apparent in this plot is that there are a lot more ship SST observations, and in fact, the ship observations are more than 1 order of magnitude greater than those

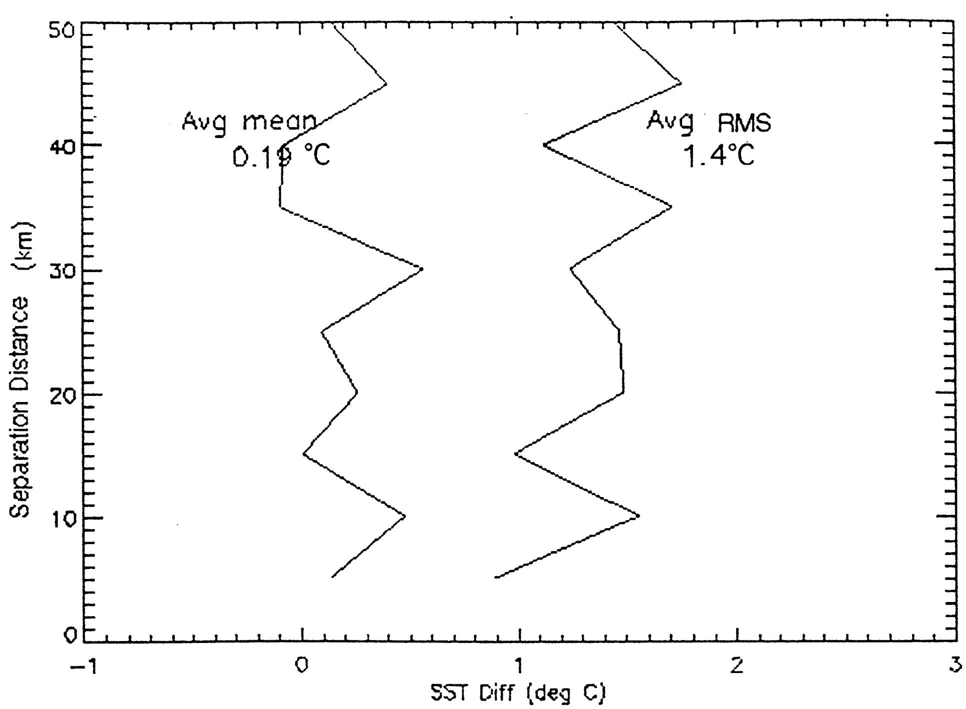

Figure 9. Ship minus drifting buoys SST for March 1996 as a function of distance

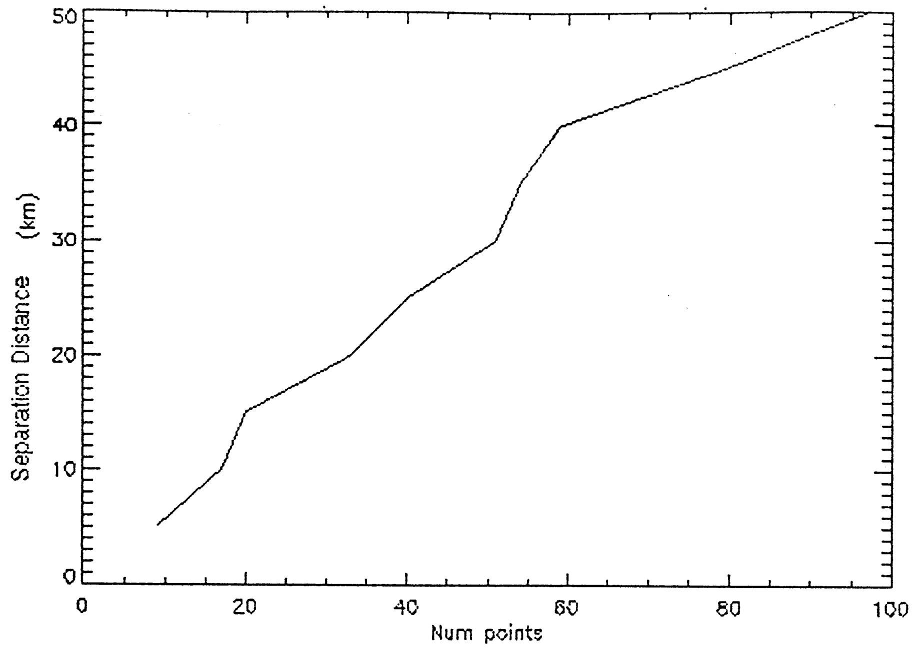

Figure 10. Number of observations of ship versus drifter SST differences for March 1996

in the corresponding drifting buoy self-comparison. As with the buoy self-comparisons this ship to ship SST comparison must have a zero mean. The RMS difference at 1.5°C is very consistent with the earlier RMS differences as it is a bit less than the ship versus buoy SST differences, and it is a bit larger than any of the comparisons between buoy SSTs.

Again, the shorter scales are expanded (Figure 12) to give more information on the short-space-scale temperature accuracies. The average RMS for the 50 km interval is 1.5°C which is about the same as the RMS difference for the overall comparison in Figure 11. For the 50 km segment the RMS difference ranges from 1.2°C at 5 km separation to 1.5°C at the 50 km distance. The number of observations (Figure 13) oscillates greatly, but all are between fairly large numbers of observations. This oscillation may reflect digitization created by the recording procedures on the ships reporting SST (due to truncation of reporting of the numbers). Note that the number of comparison points is a minimum at the 5 km, short scale and a maximum of well over 3000 at the 50 km scale. We do not understand the reason for an observation maximum at about 15 km unless this somehow represents the average distance between shipping lanes.

Seasonal variations were again investigated, and the resultant statistics are summarized in Table 3. The overall RMS temperature differences were generally close to 2.0°C but still had a range of about 0.25°C. The 50 km average RMS differences were all close to 1.5°C while the 5 km values were about 1.0°C. The number of observations was in the tens of thousands, much greater than any of our buoy comparisons.

4. Effects of In Situ Sampling Density and Geographic Bias on the SST Accuracy of the Pathfinder Algorithm

We now investigate the accuracy of the Pathfinder SST algorithm. The Pathfinder algorithm [*Podesta et al.*, 1997] is based on the assumption that the skin SSTs measured by a satellite and the bulk SSTs measured by buoys are well correlated. This is a more accurate assumption during the night when the water temperatures just below the skin tend to be constant to the depth of the buoy measurement. High winds also cause the skin and bulk SSTs to be nearly the same. However, this assumption begins to break down during daytime during regions of light winds and high solar insolation.

The Pathfinder algorithm [Kilpatrick et al., Overview of the NOAA/NASA Pathfinder algorithm for sea surface temperature and associated matchup database, 2000, manuscript in preparation] uses radiative transfer to define an equation which relates the cloud-cleared satellite radiances and the bulk SST. This equation is designed to minimize the effects of atmospheric water vapor. Historically, these equations have been modified [*Walton et al.*, 1998] when atmospheric conditions have changed (e.g., during periods of volcanic aerosols; *Reynolds*, [1993]), when new algorithms have been developed, or when new satellites have become operational. The Pathfinder algorithm uses a different tuning in different humidity regions. The equation we used can be found on the Pathfinder site available from Rosenstiel School of Marine and Atmospheric Science, University of Miami

Table 2. 1996 Ship-Drifter SST Statistics

Month	Overall RMS (°C)	50 km RMS(°C)	5 km RMS(°C)	Observed	Observed (<5 km)
Jan.	1.67	1.32	1.25	1,615	3
March	1.75	1.37	0.90	2,068	7
July	1.54	1.77	0.90	1,929	4
October	1.93	1.12	0.5	1,932	1

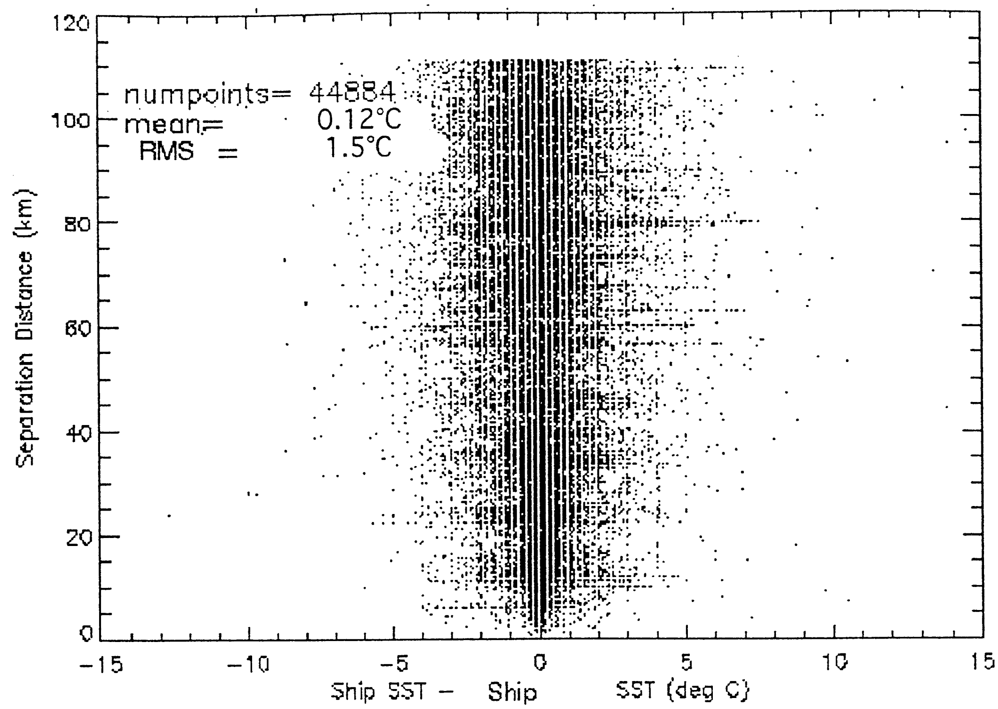

Figure 11. Ship minus ship SST differences as a function of separation distance for March 1996.

at http://www.rsmas.miami.edu/~gui/algov4/algo_updates.html, and is given by

$$SST_{sat} = a + b\,T_4 + c\,(T_4 - T_5)\,SST_{guess} + d\,(T_4 - T_5)\,[\sec(q) - 1], \quad (1)$$

where SST_{sat} is the satellite-derived SST estimate, T_4 and T_5 are brightness temperatures in the Advanced Very High Resolution Radiometer (AVHRR) channels 4 and 5, respectively, SST_{guess} is a first-guess SST value, and q is the satellite zenith angle. Coefficients a, b, c, and d are estimated from regression analyses

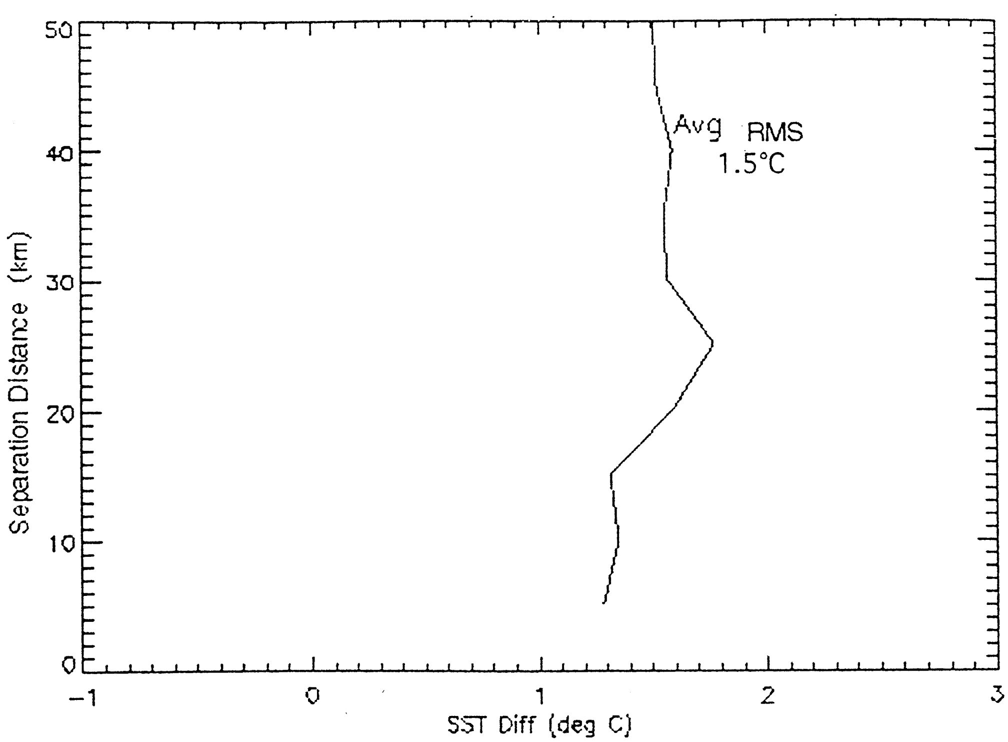

Figure 12. Ship minus ship SST for March 1996 as a function of separation distance.

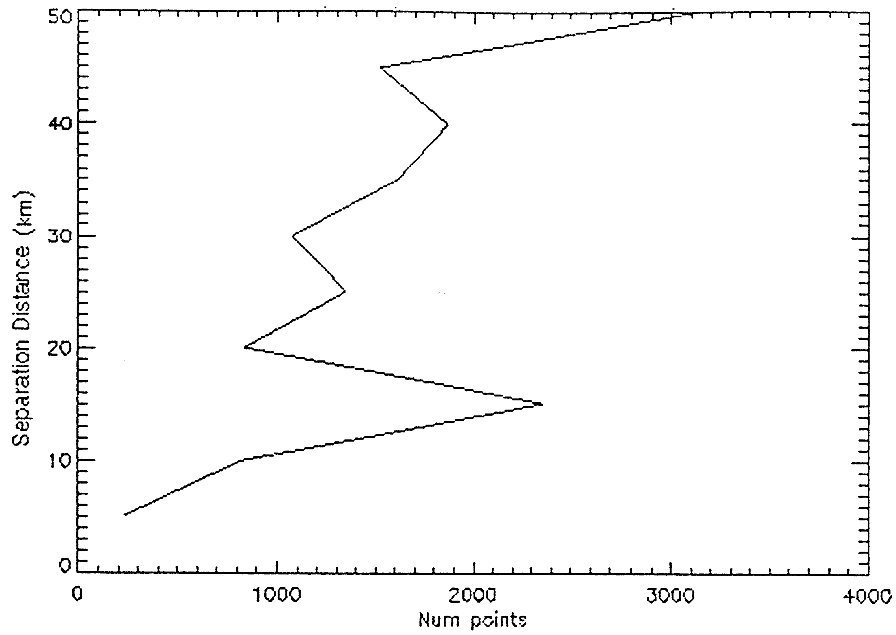

Figure 13. Number of observation pairs for ship versus ship SST comparisons for March 1999.

using collocated in situ and satellite measurements (or "matchups"). This is also the NOAA nonlinear or NLSST. Typically, NOAA produced a set of coefficients using matchups for a certain period; these coefficients would not be modified until there was a perceived need (e.g., after the eruption of the Mount Pinatubo volcano in June 1991, or when a new AVHRR was launched).

We have chosen March 1996 as a representative period to investigate the accuracy of this method. Figure 14 shows the 2461 locations of the collocated buoy and satellite observation pairs as defined by the locations of the buoy SST observations since the satellite data are basically available for at or near each drifting buoy location. These collocation were taken from the Pathfinder SST product where their definition of coincidence was within 30 min in time and 0.1° of latitude-longitude in space. The geographical distribution has the usual sampling bias to the Northern Hemisphere but otherwise appears to have a relatively good global coverage. The coefficients were determined from the collocated observations. Figure 15 shows the difference between the satellite and buoy SSTs against each buoy SST. In this case there is an insignificant residual bias with a standard deviation of 0.5°C. We refer to these statistics as dependent because all observations shown in Figure 15 were used in both the regression and the statistical error determinations. The geographic distribution has the usual sampling bias to the Northern Hemisphere but otherwise appears to have a pretty good global coverage. In this case there is an insignificant residual bias of -0.00002°C with an RMS of 0.5°C using all of the available drifting buoy SSTs.

To test the effects of sampling density on these differences, we randomly reduced the number of the calibration SSTs by 50%. We used the reduced collocated pairs (dependent) and determine our statistics using all observations (both dependent and independent). We were surprised to find that a regression using only 50% of the buoy SSTs yielded residual SST differences very similar to those found with the full in situ SST data set. The mean difference has now increased to 0.0085°C while the RMS is now only slightly larger at 0.59°C. We then sampled the buoy SSTs down to only 1% of the full data set (Figure 16) thinking that this very small population of the regression data would lead to an increase in the SST errors found by regression on these few data. However, the residual SSTs (Figure 17) were quite similar to both the 50% and 100% cases. The mean difference has now increased to 0.03°C with an RMS difference of 0.6°C. While the sampling density in Figure 17 is certainly far below that of the even the 50% case, the geographic distribution is nearly global, which is the reason that the regression SSTs are so similar to the case with full coverage. We assume that the sampling density is not as important as having a generally global distribution of regression SSTs to represent the variety of SST conditions encountered.

5. Effects of In Situ Sampling on Satellite SST Bias

To test this assumption, we examined several cases where the input regression SST observations were restricted to limited latitude bands. First, we selected buoy SSTs only from the tropics (10°S to 10°N) and computed the regression coefficients

Table 3. 1996 Ship-ship SST Statistics

Month	Overall 50 km	5 km			
	RMS (°C)	RMS (°C)	RMS (°C)	Observed	Observed (<5km)
Jan.	1.88	1.49	0.85	28,690	240
March	1.76	1.50	1.25	44,166	200
July	2.04	1.77	1.05	54,820	190
October	1.73	1.64	1.2	51,074	210

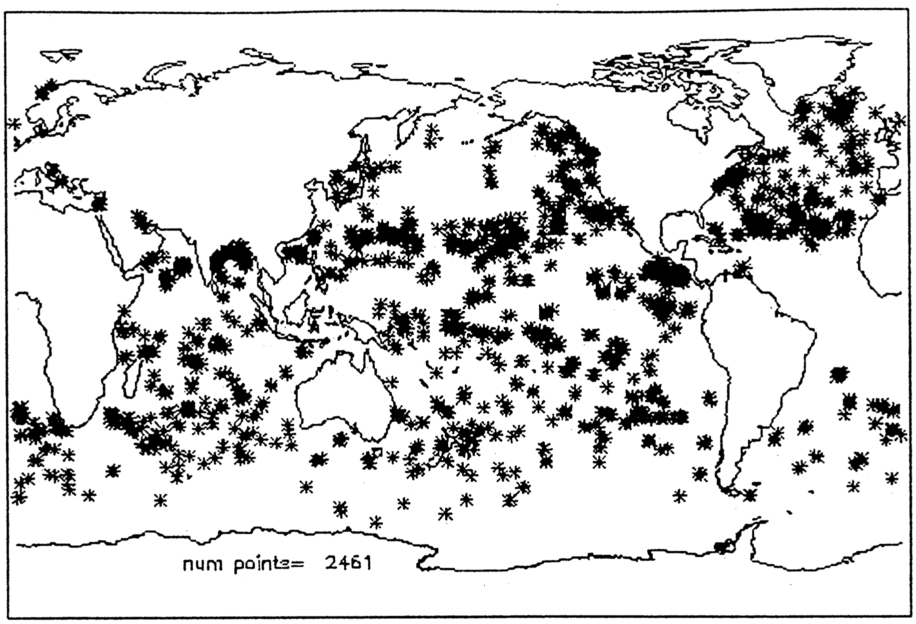

Figure 14. March 1996 geographic distribution of drifting buoys SSTs.

from these data alone. The resulting SST residuals for the full data set (Figure 18) have a rather steep increase in ΔSST at the lower buoy temperatures (where there were no SST observations in the regression parameter space). The mean residual is >1.5°C with an RMS >2.0°C. Clearly these are large errors which have been caused by the rather severe restriction in geographic coverage for the regression buoy data set.

We then altered our area restriction to using data within 30°S to 30°N, which gives a fair approximation to covering most global conditions. Turning to a polar data coverage for the regression we restricted the regression buoy SSTs to latitudes greater than 50°N and 50°S. The resulting residual SST differences (Figure 19) are quite different than any of the earlier comparisons and have a very broad scatter at the higher SST values, which is from that part of the distribution not covered by the regression data set. Thus any restriction of the regression SST observations to high or low latitudes results in marked increases in SST errors in terms of both the bias and the random variability.

To summarize this section of our study, we find that the primary requirement for an in situ calibration data set is for it to

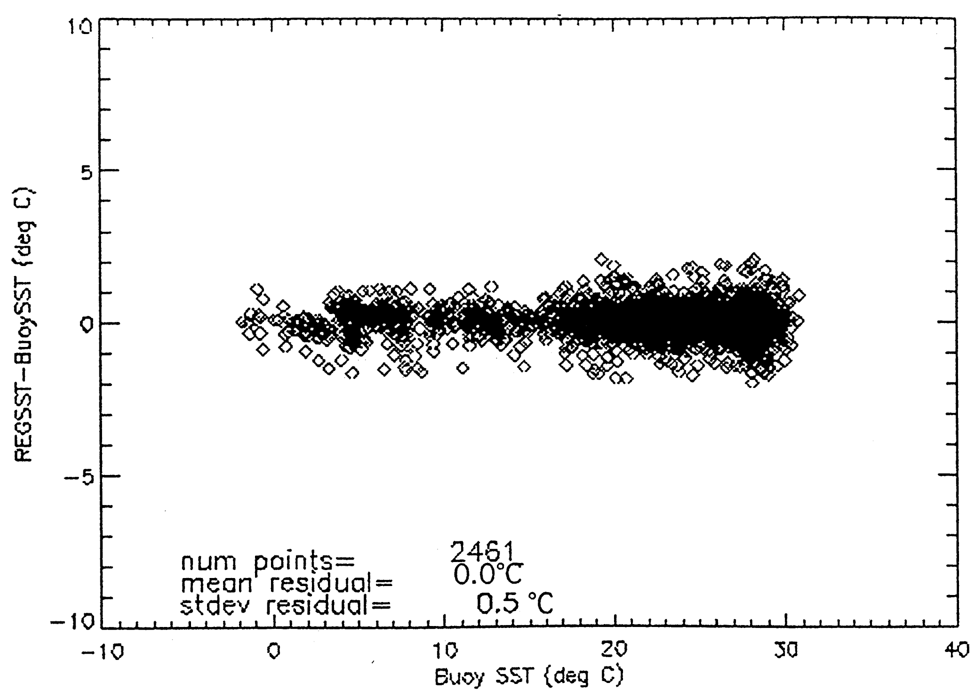

Figure 15. Satellite regression SST minus buoy SST for global data and March 1996.

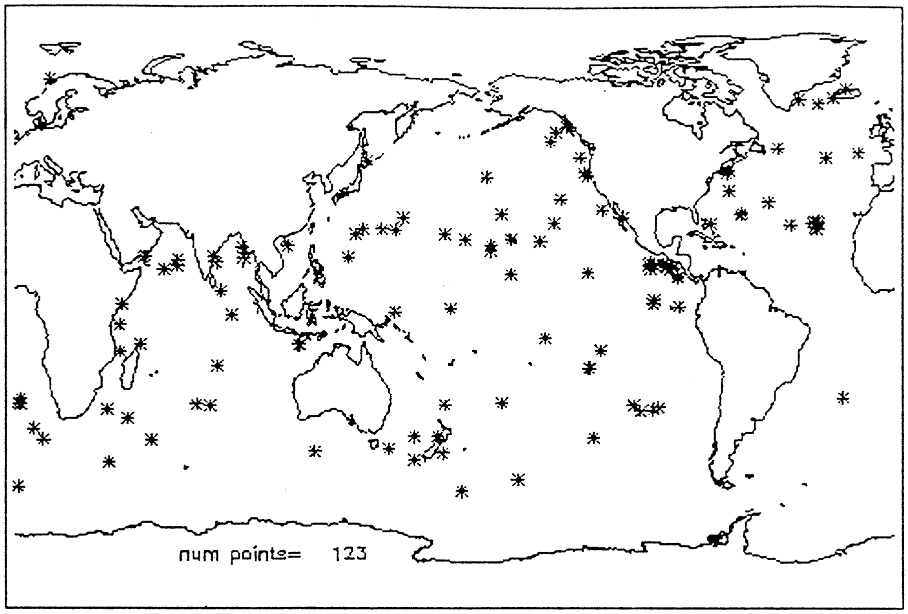

Figure 16. Geographic distribution of 1% of the March 1996 drifting buoy SSTs.

have an approximately global coverage (particularly in the meridional direction) to represent all likely SSTs that can occur. Even a fairly thin global coverage is much better than high sampling density restricted to high or low latitudes. Using data biased to either low (<10°) tropical latitudes or to polar latitudes >50° will result in large (>1.5°C) bias and large (>2.0°C) RMS SST errors. We also tried splitting the buoy regression SSTs into east and west halves for the Pacific, but the residual SSTs were about the same as the full SST case, suggesting that as expected, the east-west gradients are small when compared to the north-south gradients in their effect on SST.

The significance of this result is that as we continue to use buoy SSTs of opportunity we will be forced into using a geographic distribution that is not at all optimal for our SST calibration-validation application. At present the buoy locations are increasing at lower latitudes which effectively leaves out those conditions at higher latitudes. Other areas such as the tropical Atlantic, are being excluded, meaning that only those processes typical of the tropical Pacific will be included in the SST regression. These observations recommend an in situ buoy SST program designed strictly for the collection of SST data for satellite calibration-validation. This network of buoy SST

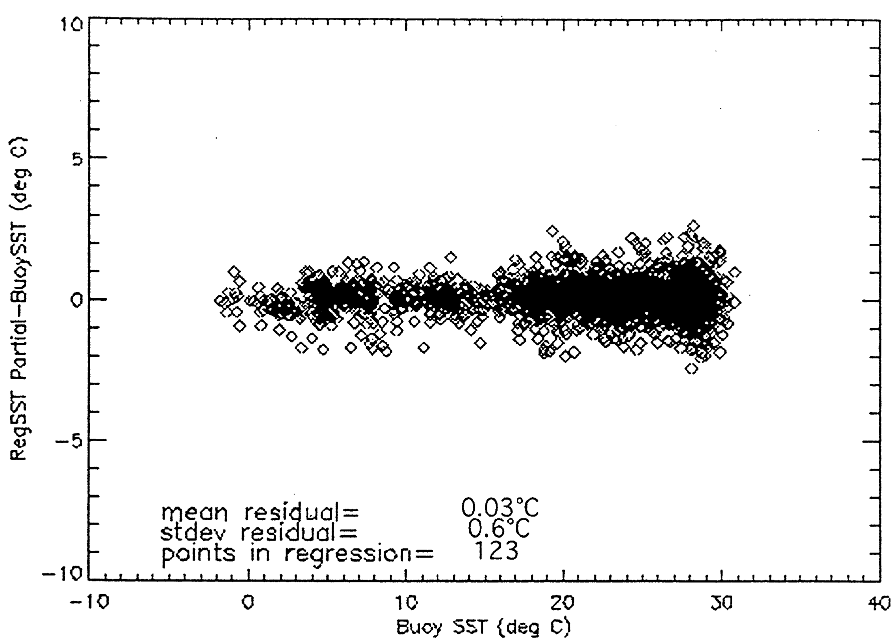

Figure 17. Regression satellite SST for 1% sample minus buoy SST for global data and March 1996.

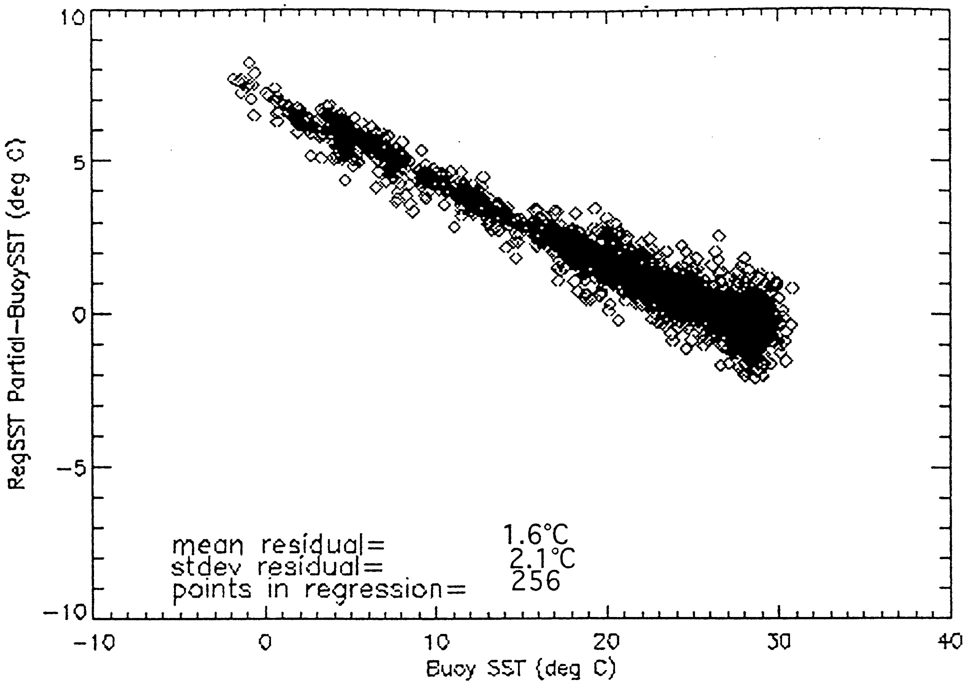

Figure 18. Regression satellite SST for tropical sample SSTs minus buoy SSTs for global data and March 1996.

measurements would include both moored and drifting buoys, all with the primary purpose of providing bulk SST data for the calibration of satellite infrared radiometers.

6. Impact of Historic Buoy Sampling Distributions on the Pathfinder SST

We wanted to evaluate the impact of "buoy of opportunity" SST measurements on the computation of the AVHRR-SST algorithm coefficients. To do this, we looked at the buoy distributions in earlier years. During the period of multichannel satellite data (November 1981 to present) the distribution and number of buoy observations have changed. We wanted to examine the impact of these changes on the satellite algorithm. For this purpose we used a cross-validation technique [e.g., see *Smith et al.*, 1996] in which the March 1996 distribution of collocated buoy and satellite observations is reduced on the

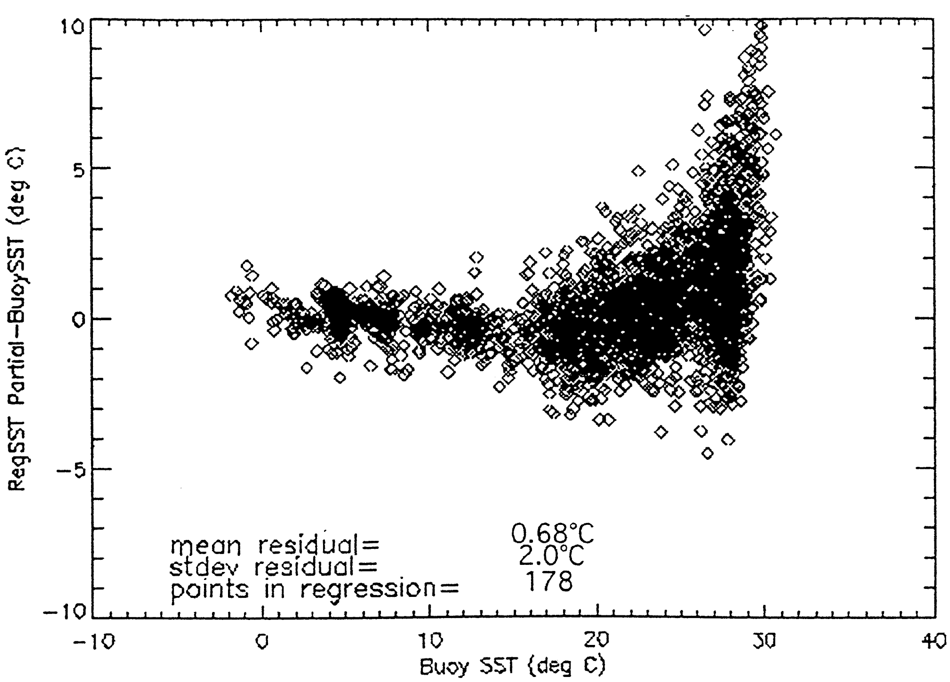

Figure 19. Regression satellite SST for greater than 50°N and for 50°S sample minus buoy SST for global data and March 1996.

basis of buoy distributions from an earlier period. This was done by computing the total number of observations on a 20° grid for the entire year of 1996 and all other years between 1982 and 1992. We then computed a ratio of the current year to 1996. If this ratio was less than 1, we reduced the number of observations for March 1996 by this ratio; if the ratio was greater than 1, the observations for March 1996 were not changed. (A ratio greater than one occurred for fewer than 5% of the total number of grid squares.) The advantage of this method is that we can simulate earlier distributions while keeping a consistent set of SST data. In the results which follow, we reduced the years discussed to a representative set: 1985, 1987, 1990, and 1992, which are shown (Figures 20a-20d). The drifting buoy distribution in March 1985 (Figure 20a) is very sparse. The Indian Ocean only has a few spots where buoy SSTs were collected. The Atlantic has a surprisingly high number in the north and far south latitudes but is completely empty in the equatorial region.

In spite of these dramatic differences in the geographic data distribution relative to 1996 (Figure 15) the summary statistics in Table 4 are surprisingly similar to the overall 1996 statistics. Note here that we have included the mean and RMS difference statistics for both March and October of each year for which sampling maps are presented for March. The mean difference is slightly (-0.13°C, -0.18°C) nonzero for both March and October but the 0.72°C RMS difference is just slightly more than the 0.6°C for 1996. As suggested by the earlier simulations it appears that the most important thing is to cover the latitude range which the 1985 distribution (Figure 20a) does in spite of its very limited meridional coverage. Here the negative mean temperature difference means that the buoy SSTs were warmer than the satellite SSTs.

This can again be seen in the March 1987 distribution (Figure 20b) which is much better filled out than that for 1985 (Figure 20a). There are now a considerable number of observations in the southern Indian Ocean, the South Pacific is surprisingly well covered, and the North Atlantic has a number of points. Poorly covered in this year is the North Pacific and the South Atlantic. The summary statistics in Table 4 are very much the same as those with the very poor coverage of 1985 with a mean of ~ -0.17°C; the RMS difference of about 0.7 is just slightly smaller than the RMS deviations for 1985. This similarity further supports the assertion that the north-south buoy SST coverage is most important in computing the satellite SST regressions.

The March distribution for 1990 (Figure 20c) is just slightly worse than that for 1987. The Indian Ocean is again almost completely empty, as is most of the North Pacific. There are some observations in the far west and east portions of the North Pacific. The South Atlantic is now empty but the North Atlantic is fairly well covered. The mean temperature difference has increased to approximately -0.24°C but the RMS difference remains about the same at 0.72°C. It is interesting that the mean difference has increased over the 1985 case, where the coverage was much worse in terms of meridional sampling.

Finally, the 1992 March distribution (Figure 20d) deviates from the earlier examples in that its coverage of both the North and South Pacific is adequate. The North Atlantic as well appears to have a good number of measurements. The Indian Ocean continues to be poorly sampled, suggesting that all of the drifting buoy projects are in the Atlantic and the Pacific. As with 1990 the Southern Ocean and the South Atlantic are rarely sampled. The difference statistics (Table 4) show that the 1992 RMS temperature differences have dropped slightly to 0.63°C for March and 0.68°C for October.

It seems clear that the practice of using drifting and moored buoys of opportunity for our satellite calibration leads to very inhomogeneous types of geographic sampling. This distribution would have a much greater impact if the important north-south variability is not accounted for. If a sampling period is well covered in north-south extent, it will give nearly the same statistics as a distribution with far greater and better distributed sampling. This is one of the reasons that the drifting buoy calibrations have worked as well as they have in the past. During this time the difference between skin and bulk SST was ignored, and the buoys were considered as "ground truth." As long as drifting buoy SST samples cover the range of conditions that occur with latitude the resulting SST algorithms will be fairly stable, and the regressions of satellite radiances onto the buoy temperatures should give a nearly correct estimate of the bulk SST.

All of the statistics in Table 4 are surprisingly similar, with mean temperature differences between -0.1° and -0.2°C,

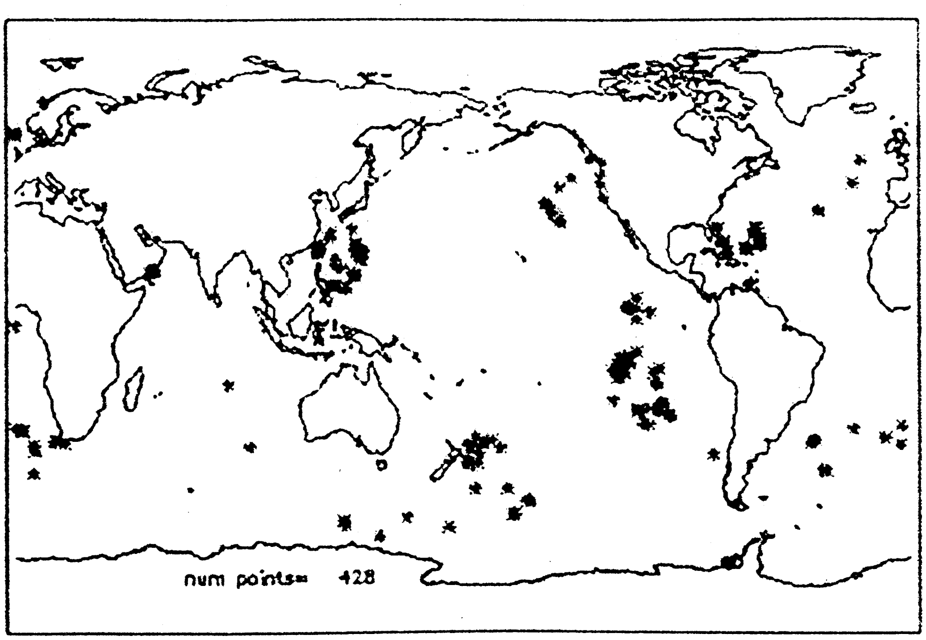

Figure 20a. March 1996 Pathfinder data sampled with March 1995 buoy distribution.

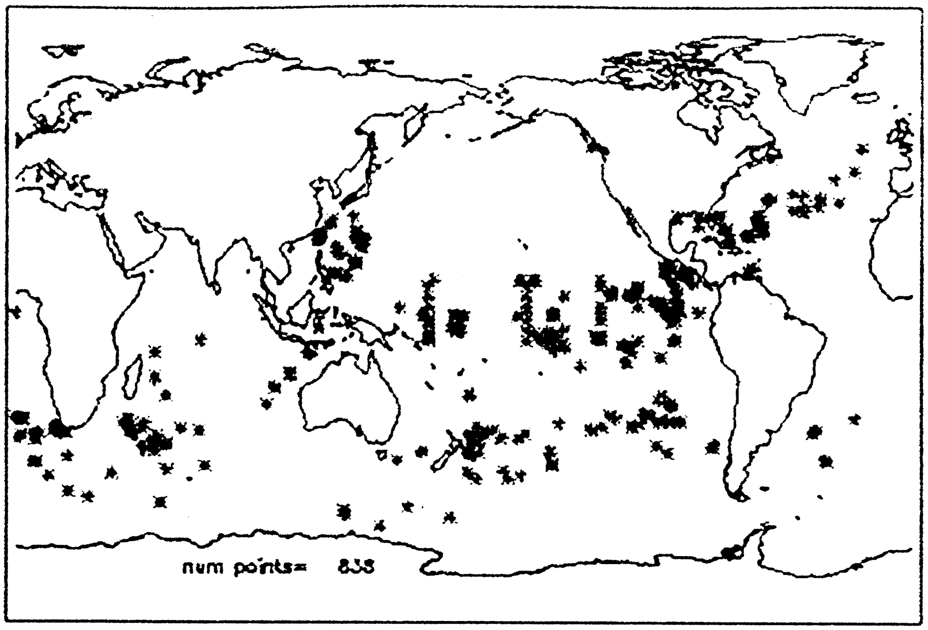

Figure 20b. March 1996 Pathfinder data sampled with March 1997 buoy distribution.

indicating that there are no strong seasonal or interannual changes in the relationships between satellite radiances and buoy SSTs. The negative biases reveal that the buoy SSTs are systematically higher than the SSTs estimated from the satellite radiances. It is tempting to attribute this bias to the skin-bulk SST effects where the "cool skin" is thought to make the satellite SSTs lower than those from the buoys. This could certainly be considered as true if we were only looking at night satellite data. In that case the skin SST must be lower (cooler) than the bulk SST. During the daytime, solar insolation can cause the skin SST to appear warmer than the 1-5 m bulk SST. The lack of seasonal and interannual temperature differences suggests that the differences we are looking at are systematic differences due to fundamental and recurring processes. While it is true that all of our measurements include both the instrument errors and geophysical variability, we have attempted to reduce this influence by looking at different times and seasons. Thus we hope that our temperature differences are more strongly related to the process and measurement errors.

The consequence of the buoy SST sampling strongly recommends that we insure the proper distribution of buoy SSTs in the future to make it possible to use the buoy SSTs with

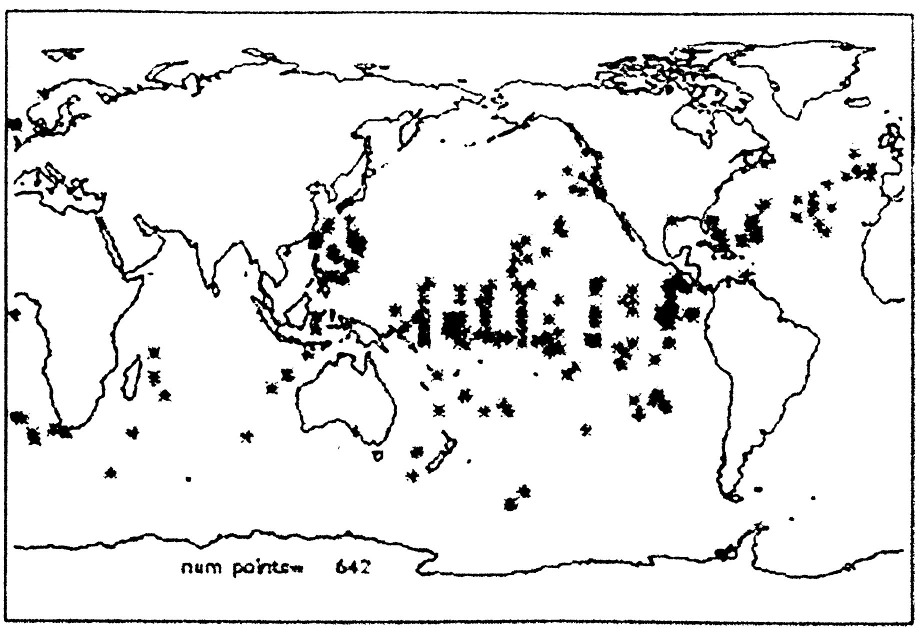

Figure 20c. March 1996 Pathfinder data sampled with March 1990 buoy distribution.

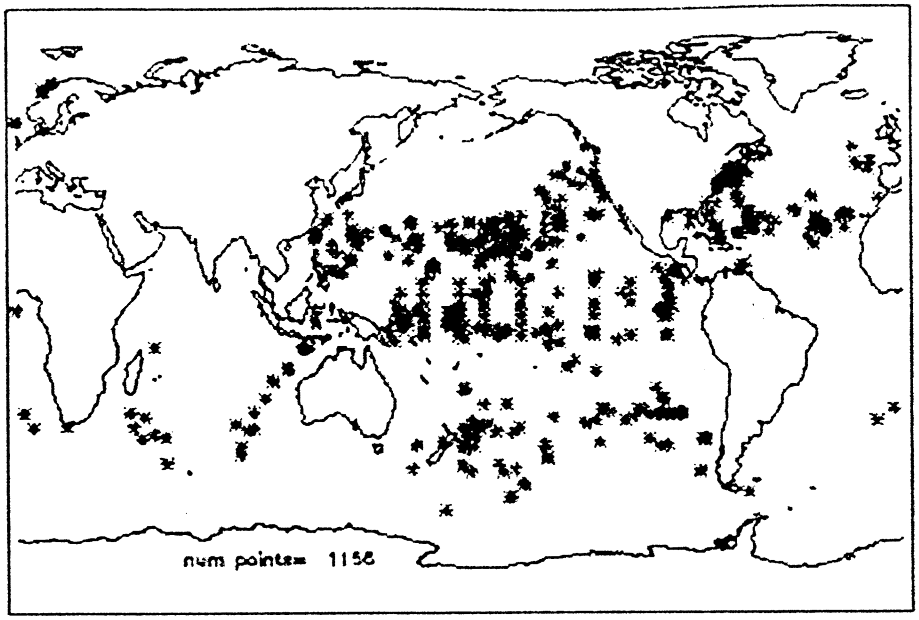

Figure 20d. March 1996 Pathfinder data sampled with March 1992 buoy distribution.

satellite radiances. The only way to do this is to have a program dedicated to providing the buoy SSTs needed for this calculation. Fortunately, moored and drifting buoy SSTs appear to behave similarly, making it possible to mix moored and buoy SSTs in the "calibration" of satellite infrared data. This means that areas like the equatorial Pacific that are well covered by moored buoys need not be sampled by drifting buoys. Other regions like the northwest Pacific, with no moored buoys, need to be sampled by drifting buoys. In light of these relationships a new program needs to be established to determine where and when drifting buoy data are needed for the calibration of the satellite infrared radiances.

7. Discussion and Conclusions

This study has explored the character of in situ measurements of bulk SST as reported by drifting-moored buoys and merchant vessels. The general geographic distribution of the buoy SSTs shows a very clear bias to warm temperatures collected at tropical and subtropical latitudes. Individual years such as 1990 have an even worse geographic distribution with coverage only in the tropical Pacific and high-latitude North Atlantic. There are very few drifting buoy SSTs being reported from polar or subpolar latitudes. Other places missing for the 1996 study period are the equatorial Atlantic/South Atlantic, the Indian Ocean, and the western upwelling regions of the Americas. This means that any satellite SST algorithm computed from these drifting buoy populations will be biased toward the warm temperatures of the lower latitudes leading to some substantial errors in SSTs estimated for higher latitudes. This condition is much more severe in some years than in others, and any algorithm based on global drifting buoy observations must take this distribution into account.

The inherent accuracy of these buoy and ship measurements was explored using a unique computation of temperature (SST) differences as a function of separation distance for buoys or ships reporting within the same hour. For drifting buoy data and separation intervals between 0 and 50 km the mean difference is ~ 0.05°C with an RMS of ~ 0.4°C. This suggests that the buoy SST are very consistent but have a basic variability that results in an RMS difference of about 0.4°C. This variability number includes natural variability of the local SST field, calibration and other sensor errors. A large number of the uninstalled buoy SST sensors are calibrated to ± 0.1°C but experience with the SST data (M. Swenson, personal communication) suggests that an accuracy of ± 0.15°C is more consistent with the buoy SST data collected. Unfortunately, there are no calibration histories for these drifters, which are considered expendable and are not retrieved for a postoperation calibration. As a result we must infer the accuracies of the drifting buoy SSTs from the data collected. As a consequence we must combine both geophysical variability with instrument noise. The assumption here is that as we reduce the separation between buoy SSTs, we gradually eliminate the geophysical contribution to the SST RMS variability and attain an estimate of the basic measurement noise. A comparison between drifter SSTs at the shorter separations has a global variability of 0.4°C. Comparing drifting buoy SSTs with

Table 4. March and October Satellite Regression Statistics for Simulated Distributions Relative to 1996

Distribution Year	Mean Temperature Difference °C		RMS Temperature Difference °C	
	March	Oct.	March	Oct.
1985	-0.13	-0.18	0.73	0.72
1987	-0.18	-0.17	0.70	0.69
1990	-0.23	-0.24	0.71	0.72
1992	-0.09	-0.24	0.63	0.68

coincident moored buoy SSTs for the extratropics, we find that there is a slight bias of -0.1°C indicating that the moored SST is cooler than that for the drifting buoys. The same comparison carried out for the equatorial Pacific separately gave a mean of 0.05°C and an RMS difference of 0.1°C. The statistics for the ship SSTs result in significantly larger errors (mean of about -0.15°C and RMS difference of ~ 1.2 °C) as would be expected from the less homogeneous ship temperature sensors most of which are not calibrated and whose analog measurements are not regularly checked for calibration.

We need to realize that most, if not all, of the drifters are deployed to study some phenomenon other than to measure and monitor SST for satellite radiometer calibration-validation. While this latter observational aspect is frequently invoked as added justification for the buoy deployments, there is generally some other specific goal for the buoy deployments. This can easily be seen in the data distribution for 1990, when there was a clear emphasis on the tropical Pacific and the subpolar North Atlantic. For this reason we have a global drifting buoy SST distribution that emphasizes the lower latitudes and is mostly unsampled at higher latitudes. In certain regions, such as the equatorial and subtropical South Atlantic, for the period studied, there are no buoy SST observations reported. We need to also recognize that the ultimate distribution of the buoy SSTs will be dictated by the ocean currents and, while that information is useful, it does not help us with our satellite SST calibration-validation issues.

It is also important to realize that the meridional coverage is much more important than the zonal sampling. Essentially the same results were given very different overall distributions that had the same overall meridional coverage. When the buoy SST coverage was restricted by latitude, very large (on the order of 2°C) regression errors were introduced. The realization that the meridional coverage is much more important than the zonal distribution might actually make it easier to collect adequate buoy SST samples. This meridional coverage can be collected by buoys deployed north and south along the east and west coastlines rather than in the less accessible open ocean. It is also important to recognize that in the future it will be necessary to design and deploy SST measuring buoys to calibrate and validate the satellite SST measurements. It is not a good idea to continue to rely on the "platform of opportunity principle" that relies on other measurement programs to carry out their data collection. Only by establishing an SST measurement program dedicated to providing data for satellite calibration can we be sure that we will have the data we need for this application. No amount of processing can overcome the spatial and temporal sampling biases that exist in a buoy of opportunity in situ SST measurement program.

Acknowledgements

This research was supported by two different grants from the Earth Science Enterprise of the National Aeronautics and Space Administration (NASA). The contributions of the program managers (Jim Dodge and Eric Lindstrom) are greatly appreciated.

References

Evans, R., and G. Podesta, AVHRR Pathfinder SST approach and results, *EOS trans*, Fall Meeting Supple., F354, 1996.

Kent, E.C., and P.K. Taylor, Choice of a Beaufort equivalent scale, *J. Atmos. Oceanic Technology., 14*, 228-242, 1997.

Kent, E. C., P. K. Taylor, B. S. Truscott and J. A. Hopkins, The accuracy of voluntary observing ship's meteorological observations, *J. Atmos. Oceanic Technology*, ***10***, 591-608, 1993.

Kent, E.C., P.G. Challenor and P.K. Taylor, A Statistical determination of the random observational errors present in voluntary observing ships meteorological reports, *J Atmos. & Oceanic Technology, 16*, 905-914, 1999.

McClain, E.P., W. G. Pichel, C.C. Walton, Z. Ahmad, and Sutton, Multi-channel improvements to satellite-derived global sea surface temperatures, *Adv. Space Res., 2*, 43-47, 1983.

McClain, E.P., W.G. Pichel, and C.C. Walton, Comparative performance of AVHRR-based multichannel sea surface temperatures, *J. Geophys. Res., 90*, 11,587-11,601, 1985.

Podesta, G. P., S. Shenoi, J. W. Brown, and R. H. Evans, AVHRR Pathfinder Oceans Matchup Database, *Tech. Rep., University of Miami 069_D001* Miami, Fla,. 1997.

Reynolds, R.W., Impact of Mount Pinatubo aerosols on satellite-derived sea surface temperature, *J. Climate, 6*, 768-774, 1993.

Saur, J.F.T., A study of the quality of sea water temperatures reported in logs of ships weather observations, *J. Appl. Meteorolo., 2*, 417-425, 1963.

Slutz, R.J., S.J. Lubker, J.D. Hiscox, S.D. Woodruff, R.L. Jenne, D.H. Joseph, P.M. Steuer, and J.D. Elms, Comprehensive ocean-atmosphere data set; release 1, ., 268 pp, NOAA Environ. Res. Lab., Climate Res. Program, Boulder, Co, 1985.

Smith, T.M., R.W. Reynolds, R.E. Livezey, and D.C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions, J. Climate, 9, 1,403-1,420, 1996.

Tabata, S., On the accuracy of sea-surface temperatures and salinities observed in the northeast Pacific Ocean, *Atmos - Ocean, 16*, 237-247, 1978.

Walton, C. C., W.G. Pichel, J. F. Sapper and D. A. May, The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA Polar-Orbiting Environmental Satellites, *J. Geophys. Res.*, 103, 27,999-28,012, 1998.

D. J. Baldwin and W. J. Emery, Engineering Center, Aerospace Engineering Sciences Department, CCAR Box 431, University of Colorado, Boulder, Co., 80309 (emery@frodo.colorado.edu)

R. W. Reynolds, NCDC-NESDIS-NOAA, 5200 Auth Road #807, Camp Springs, MD, 20746. (rreynolds@ncep.noaa.gov)

P. Schluessel, EUMETSAT, Am Kavalleriesand 31, 64295 Darmstadt, Germany (schluessel@eumetsat.de)

(Received January 31, 2000; revised August 3, 2000; accepted August 14, 2000.)